

# Paul Lévy, strong approximation and the St. Petersburg paradox

István Berkes \*

## Abstract

The St. Petersburg paradox, formulated in the early 1700's, concerns the 'fair' entry fee in a game where the winnings are distributed as  $P(X = 2^k) = 2^{-k}$ ,  $k = 1, 2, \dots$ . Since the tails of  $X$  are not regularly varying, the accumulated gain  $S_n$  in  $n$  St. Petersburg games has no limit distribution after any centering and norming, making the asymptotic study of the game a challenging problem. The problem was solved by Martin-Löf (1985) and Csörgő and Dodunekova (1991), leading to a clarification of the paradox and a fascinating asymptotic theory. The purpose of this paper is to discuss the crucial, but forgotten contribution of Paul Lévy (1935) to the field. In a remark in his classical paper [23], Lévy determines the asymptotic distribution of a large class of i.i.d. sums with generalized St. Petersburg tails  $cx^{-\alpha}\psi(\log x)$ , where  $0 < \alpha < 2$  and  $\psi$  is a periodic function on  $\mathbb{R}$ . His proof uses a coupling argument similar to Skorohod representation and provides a strong (pointwise) approximation result, the first in probability theory. The argument also yields a strong approximation approach to domains of attraction and partial attraction, as well as strong approximation of i.i.d. sums with infinitely divisible variables.

We finally discuss an argument of Lévy [23], also of considerable historical interest, proving a limit theorem via the quantile transform, another 'first' in probability theory. His argument yields a qualitative version of the stable decomposition theorem of LePage, Woodroffe and Zinn (1981) and, adapted to semistable variables, leads to a complete solution of the strong approximation problem for St. Petersburg sums, as we will show in a subsequent paper [3].

## 1 Introduction

Let  $X, X_1, X_2, \dots$  be i.i.d. r.v.'s with

$$P(X = 2^k) = 2^{-k}, \quad (k = 1, 2, \dots) \quad (1)$$

---

\*Graz University of Technology, Institute of Statistics, Kopernikusgasse 24, 8010 Graz, Austria. e-mail: [berkes@tugraz.at](mailto:berkes@tugraz.at). Research supported by FWF grant P24302-N18 and NKFIH grant K 108615.

and let  $S_n = \sum_{k=1}^n X_k$ . The asymptotic behavior of the sequence  $\{S_n, n \geq 1\}$  has attracted considerable attention in the literature in connection with the St. Petersburg paradox (for the 'standard' formulation, see Daniel Bernoulli [5]), concerning the 'fair' entry fee in a game where the winnings are distributed as  $X$ . We refer to Csörgő and Simons [11] for a historical account and bibliography of the problem. Solving the entry fee problem requires determining the precise asymptotic behavior of  $S_n$ . Feller [14] proved that

$$\lim_{n \rightarrow \infty} \frac{S_n}{n \log_2 n} = 1 \quad \text{in probability} \quad (2)$$

(where  $\log_2$  denotes logarithm with base 2) and Martin-Löf [25] obtained

$$S_{2^k}/2^k - k \xrightarrow{d} G,$$

where  $G$  is the infinitely divisible distribution function with characteristic function  $\exp(g(t))$ , where

$$g(t) = \sum_{l=-\infty}^0 (e^{it2^l} - 1 - it2^l)2^{-l} + \sum_{l=1}^{\infty} (e^{it2^l} - 1)2^{-l}. \quad (3)$$

He also proved that if  $n_k \sim \gamma 2^k$ ,  $1 \leq \gamma < 2$ , then

$$S_{n_k}/n_k - \log_2 n_k \xrightarrow{d} G_\gamma \quad (4)$$

where  $G_\gamma$  denotes the distribution with characteristic function  $\exp(\gamma g(t/\gamma) - it \log_2 \gamma)$ . Letting  $\gamma_n = n/2^{\lfloor \log_2 n \rfloor}$  (where  $\lfloor \cdot \rfloor$  denotes integral part), Csörgő and Dodunekova [9] proved that (4) holds iff  $\gamma_{n_k} \rightarrow \gamma$  and Csörgő ([8], Theorem 1) proved that

$$\sup_x \left| P\left(\frac{S_n}{n} - \log_2 n \leq x\right) - G_{\gamma_n}(x) \right| \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad (5)$$

and determined the precise convergence rate. Relation (5) shows that the class of subsequential limit distributions of  $S_n/n - \log_2 n$  is the class

$$\mathcal{G} = \{G_\gamma : 1 \leq \gamma < 2\}.$$

If  $n$  runs through the interval  $[2^k, 2^{k+1}]$ , then  $G_{\gamma_n}$  moves through the distributions  $G_{j/2^k}$ ,  $2^k \leq j \leq 2^{k+1}$  representing, in view of  $G_1 = G_2$  (cf. [25], Theorem 2), a "circular" path in  $\mathcal{G}$ . In view of (5), the distribution of  $S_n/n - \log_2 n$  also describes approximately a circular path, a remarkable asymptotic behavior called *merging* in [8]. Using strong approximation of the uniform empirical process, Csörgő and Dodunekova [9] showed that merging holds for extremal and trimmed sums of the sequence  $(X_n)$  as well and Berkes, Horváth and Schauer [4] and del Barrio, Janssen and Pauly [2] proved that the same holds for bootstrapped sums.

The purpose of this paper is to discuss the crucial, but forgotten contribution of Paul Lévy to the field. In Section 2 we will discuss a remark in Lévy [23] stating and proving the basic limit theorem for St. Petersburg sums for a much larger class of i.i.d. sequences. Lévy's proof, depending on his construction of semistable laws and a remarkable coupling argument, provides a particularly simple and elementary approach to the problem and fills a crucial missing piece in St. Petersburg history. His method works for general i.i.d. sums as well, providing a strong approximation approach to domains of attraction and domains of partial attraction, as well as approximation of i.i.d. sums with infinitely divisible variables. Finally, we will discuss an argument in [23] concerning stable distributions which appears to be the first application of the quantile transform method in probability theory and which, adapted to semistable variables, leads to a complete solution of the strong approximation problem for St. Petersburg sums, as we will show in a subsequent paper [3].

## 2 Asymptotics by Poisson coupling

Let  $X, X_1, X_2, \dots$  be i.i.d. random variables and let  $S_n = \sum_{k=1}^n X_k$ . If for some numerical sequences  $(a_n), (b_n)$  we have

$$(S_n - a_n)/b_n \xrightarrow{d} Z \quad (6)$$

with a nondegenerate  $Z$ , then  $Z$  is either Gaussian or  $\alpha$ -stable with some  $0 < \alpha < 2$ . A necessary and sufficient criterion for a Gaussian limit is

$$\lim_{x \rightarrow \infty} \frac{x^2 P(|X| > x)}{EX^2 I\{|X| \leq x\}} = 0 \quad (7)$$

and the corresponding criterion for an  $\alpha$ -stable limit is that  $P(|X| > x)$  is regularly varying with exponent  $-\alpha$  and

$$\lim_{x \rightarrow \infty} \frac{P(X > x)}{P(|X| > x)} = p, \quad \lim_{x \rightarrow \infty} \frac{P(X \leq -x)}{P(|X| > x)} = q \quad (8)$$

for some  $p, q \geq 0, p + q = 1$ . The criterion in the Gaussian case is due to Lévy ([23], Théorème II, p. 366 and [24], Théorème 36,3, p. 113); the stable case was obtained independently by Gnedenko [15] and Doeblin [13]. Note that their criterion was anticipated by Lévy, who in [23], p. 374, remarks that

*I would like to stress here the profound difference between convergence to the Gaussian law [...] and convergence to other stable laws, which is the consequence of precise hypotheses on the probability of large values of the variable. One gets one of these stable laws as a limit only if the original law resembles that sufficiently. [...]*

Then he goes on to say that

*There is an analogous difference between the convergence to a Gaussian law in the case when  $\mathcal{E}\{x^2\}$  is infinite and the case of convergence to another stable law. Assume that all the  $x_n$  have (except perhaps in a finite interval) the same distribution which we assume, to fix the ideas, to be symmetric and such that  $F(X) = X^{-\alpha}$  ( $0 < \alpha \leq 2$ ). Divide the interval  $(1, \infty)$  into infinitely many intervals separated by the numbers  $X_n = q^n$  ( $q > 1$ ), and in each of in these intervals perform an arbitrary change of the probability distribution. If  $\alpha = 2$ , these modifications have no effect on the type of the limit distribution, which is Gaussian. [...] If, on the contrary,  $\alpha < 2$ , then the modifications in the different intervals cause perturbations which act successively when  $n$  grows and for each of them there comes a moment when its effect is not any more negligible.[...] If the modifications occur periodically, we will have convergence to a class of associated semistable laws (here we mean a semistable law and its powers whose characteristic functions are the powers of the original characteristic function).*

A central issue in [23] is the connection between the behavior of a distribution function at  $\pm\infty$  and the asymptotic behavior of the partial sums of the corresponding i.i.d. sequence. In addition to the main, and now classical, theorems of the paper, Lévy also discusses many interesting irregular situations, one of which is the above example. He gives no explicit direct proof of his claim, but a proof can be recovered from his construction of semistable laws and a remarkable coupling argument in the case of stable attraction. The symmetry of the  $x_k$  is not used in the proof; what Lévy's argument actually yields is that if  $0 < \alpha < 2$ ,  $q \geq 2$  is an integer and  $x_n$  are i.i.d. random variables with tails of the form

$$P(x_n > x) = c_1 x^{-\alpha} \psi(\log_q x), \quad P(x_n \leq -x) = c_2 x^{-\alpha} \psi(\log_q x) \quad (x \geq x_0) \quad (9)$$

where  $\psi$  is a bounded function with period  $1/\alpha$ , then the class of weak subsequential limits of  $n^{-1/\alpha} \sum_{k=1}^n x_k$ , suitably centered, is the class of convolution powers of the infinitely divisible semistable distribution whose Poisson measure has the tails in (9) for  $x > 0$ . The periodic tail condition (9), formulated in Lévy's remark only verbally, is a consequence of the structure of semistable laws, see p. 357, 3<sup>o</sup>. Except a minor technical point (see Footnote 1), Lévy's proof is complete in the case  $\alpha \neq 1$ ; in the case  $\alpha = 1$  his argument is sketchy, in particular, he does not determine the centering factor in the limit theorem. This gap can be removed easily by truncation, leading to the centering factor  $c \log n$ .

Lévy's argument, written in his characteristic style, is a most interesting reading both for probabilists and anyone interested in the history of probability theory. Below we give his proof, using today's terminology, in the case of positive variables  $x_n$ . Assume first  $0 < \alpha < 1$ , when no centering factor is needed and the argument is the simplest. In [23], pp. 352–356 Lévy gives his now classical construction of stable laws by inhomogeneous Poisson processes and on pp. 358–359 he uses it to prove that if  $x_n$  are i.i.d. positive random variables with tails  $P(x_n > x) = x^{-\alpha}$  for  $x \geq 1$ ,

then  $n^{-1/\alpha} \sum_{k=1}^n x_k$  converges weakly to a completely asymmetric stable law  $\mathcal{L}$  with parameter  $\alpha$ . (Of course, this holds under more general conditions and is easily proved by characteristic functions, but Lévy's direct coupling argument has great historical interest.) Let  $y_1 > y_2 > \dots$  be the points of an inhomogeneous Poisson process  $\mathcal{P}$  on  $(0, \infty)$  with intensity measure having tails

$$T(x) = x^{-\alpha} \quad (x > 0). \quad (10)$$

Let  $S = \sum y_i$ , and let  $\mathcal{L}$  denote the distribution of  $S$ . The relation  $S < \infty$  a.s. follows from the fact that the expected number of points  $y_i$  in  $(1, \infty)$  is  $T(1) < \infty$  and  $E \sum y_i I(y_i \leq 1) = \int_0^1 xT(dx) < \infty$  by  $\alpha < 1$ . Since  $nT(n^{1/\alpha}x) = T(x)$  for any  $n \geq 1$ , the sum of  $n$  random variables with distribution  $\mathcal{L}$ , divided by  $n^{1/\alpha}$  has again distribution  $\mathcal{L}$  and thus  $\mathcal{L}$  is stable. (In accordance with his earlier book [22], Lévy uses the notation  $\mathbf{L}_{\alpha, -1}$  for this distribution, which is the  $\alpha$ -stable law with skewness parameter  $\beta = -1$ , see also Gnedenko and Kolmogorov [17], p. 164). Clearly,

$$\mathcal{L} = \lim_{t \rightarrow 0} \mathcal{L}^{(t)}$$

in law, where  $\mathcal{L}^{(t)}$ ,  $t > 0$  is the distribution of the sum  $\bar{S}_t$  of the points of the Poisson process  $\mathcal{P}$  exceeding  $t$ . The number of such points is Poisson distributed with mean  $T(t)$  and the distribution of  $\bar{S}_t$  equals the distribution of the sum of  $N$  i.i.d. random variables concentrated on  $(t, \infty)$  with tails proportional to  $T(x)$ , where  $N$  is a Poisson variable having mean  $T(t)$  and independent of the i.i.d. sequence. Let now  $x_1, x_2, \dots$  be i.i.d. random variables concentrated on  $(1, \infty)$  with tails  $T(x)$  and  $\hat{S} = n^{-1/\alpha} \sum_{k=1}^n x_k$ . Clearly,  $\hat{S}$  is obtained from  $\bar{S}_t$  by choosing  $t = n^{-1/\alpha}$  and replacing the Poisson variable  $N$  by  $n$ . Since the mean and variance of  $N$  is  $T(n^{-1/\alpha}) = n$ , Chebysev's inequality yields  $|N - n| = O_P(\sqrt{n})$ , which implies easily<sup>1</sup> that the Lévy distance between the distributions of  $\bar{S}_t$  for  $t = n^{-1/\alpha}$  and  $\hat{S}$  tends to 0 as  $n \rightarrow \infty$  and, consequently,  $\hat{S} \xrightarrow{d} \mathcal{L}$  as  $n \rightarrow \infty$ , i.e.  $n^{-1/\alpha} \sum_{k=1}^n x_k \xrightarrow{d} \mathcal{L}$ .

In the case  $1 < \alpha < 2$  the argument is similar, only from the  $x_\nu$  one has to subtract their means  $Ex_\nu$  and the definition of  $S$  has to be replaced by  $S = \lim_{t \rightarrow 0} (\bar{S}_t - E\bar{S}_t)$ . Here  $E\bar{S}_t = \int_t^\infty xT(dx) < \infty$  by  $\alpha > 1$  and the existence and a.s. finiteness of  $S$  follows, as Lévy remarks, from the Kolmogorov two series criterion. More changes are needed in the case  $\alpha = 1$  and Lévy only sketches them; in particular, he does not compute the centering factor in the limit theorem for  $\sum_{k=1}^n x_k$ . In the stable case  $\psi = \text{const}$  he computes the value  $c_n = c \log n$  in [24], p. 209; the argument there works in the semistable case as well.

Let now  $\psi(x)$  be a bounded periodic function on  $\mathbb{R}$  with period  $1/\alpha$  and assume that

$$T(x) = x^{-\alpha} \psi(\log_q x), \quad x > 0 \quad (11)$$

---

<sup>1</sup>This requires a probability estimate for the maximal fluctuation of the partial sum process  $\{S_j, j \geq 1\}$  of the sequence  $(x_k)$  for  $n - Cn^{1/2} \leq j \leq n + Cn^{1/2}$ , which follows from maximal inequalities known at the time.

is nonincreasing. Then, as Lévy points out (see [23], p. 357, 3<sup>o</sup>, "On ne retrouve...") for the special values  $n = q^k$  we still have  $nT(n^{1/\alpha}x) = T(x)$  and thus replacing the Poisson process corresponding to (10) by the one corresponding to (11) and denoting the distribution of  $S$  again by  $\mathcal{L}$ , the sum of  $n = q^k$  random variables with distribution  $\mathcal{L}$ , divided by  $n^{1/\alpha}$  has distribution  $\mathcal{L}$  and thus  $\mathcal{L}$  is semistable. Also, the previous argument yields  $\hat{S} \xrightarrow{d} \mathcal{L}$  as  $n \rightarrow \infty$  along these special  $n$ 's, proving Lévy's claim for the indices  $n = q^k$ . To prove the general case, in [23], p. 380 Lévy points out that

*Indeed, if for a sequence  $n_p$  of  $n$ 's one gets laws whose types converge to that of a law  $\mathcal{L}'$ , then for values  $n'_p$  such that  $n'_p/n_p$  has a limit  $k$ , one gets laws whose types converge to that of  $\mathcal{L}'^k$ , which denotes the law whose characteristic function is obtained by raising that of  $\mathcal{L}'$  to the power  $k$ .*

Note that this principle holds in full generality: if along a subsequence  $(n_p)$  the centered and normed partial sums of an i.i.d. sequence converge weakly to a distribution  $G$ , then along another subsequence  $(n'_p)$  with  $n'_p/n_p \rightarrow k$  where  $k$  is an arbitrary positive number, the suitably centered and normed partial sums converge weakly to the convolution power  $G^{*k}$  (that is, to the law with characteristic function  $\varphi^k$ , where  $\varphi$  is the characteristic function of  $G$ ). The proof (which Lévy omits) is immediate by characteristic functions. Since  $\varphi^k$  is the pointwise limit of a sequence of characteristic functions, it is itself a characteristic function. Thus in the case of the sequence  $(x_n)$  in (9), the centered and normed partial sums with indices  $n_k \sim cq^k$ ,  $1 \leq c < q$ , converge weakly to the  $c$ -th convolution power of the limit distribution for  $c = 1$ .

Lévy's remark quoted above is repeated, without proof, by Doeblin [13] (see the paragraph after Théorème II on p. 78) and with proof by Gnedenko ([16], Theorem 3). Gnedenko uses this fact to prove that if a distribution  $G$  is not stable, then the class  $\mathcal{G}$  of different (modulo linear transformations) convolution powers  $G^{*c}$  is uncountable (see also Gnedenko and Kolmogorov [17], p. 189). Csörgő ([7], Theorem 10) proved that  $\mathcal{G}$  actually has the cardinality of the continuum.

Note that Lévy's proof of the limit relation  $\hat{S} \xrightarrow{d} \mathcal{L}$  above, which expresses his claim for  $n = q^k$ , is a coupling argument, defining the sequence  $(x_n)$  and the Poisson r.v.  $N$  on the same probability space and comparing pointwise the sequences  $\hat{S} = n^{-1/\alpha} \sum_{k=1}^n x_k$  and  $\tilde{S}_t = n^{-1/\alpha} \sum_{k=1}^N x_k$  ( $t = n^{-1/\alpha}$ ), the latter of which converges weakly to  $\mathcal{L}$ . Note the similarity of this method to Skorohod embedding: Lévy's method approximates St. Petersburg sums by randomized i.i.d. sums, while Skorohod embedding represents partial sums of square integrable i.i.d. sequences as a randomly stopped Wiener process. In a sense, Lévy's idea is complementary to Skorohod's: it represents not the partial sums of the  $x_\nu$  themselves, but, after a small perturbation, the limiting semistable variable. The method also has important consequences for general i.i.d. sums. Let  $x_n$  be i.i.d. random variables concentrated on  $[1, \infty)$  with tails  $G(x) = P(x_1 > x)$ , let  $S_n = \sum_{k=1}^n x_k$ , let  $N$  be a Poisson random variable with mean  $n$ , independent of the  $x_k$ 's, and let  $(a_n)$  be a positive numerical sequence. Clearly,  $S_N/a_n$  has the same distribution as the sum of points in a nonhomogeneous

Poisson process  $\mathcal{P}$  in  $(1/a_n, \infty)$  with intensity measure with tails  $T(x) = nG(a_n x)$ . This distribution is infinitely divisible with Lévy measure concentrated on  $(1/a_n, \infty)$  with tails  $nG(a_n x)$ , and thus, apart from the fact that the support of the Lévy measure is  $(1/a_n, \infty)$  instead of  $(0, \infty)$ , it is the accompanying infinitely divisible distribution to  $S_n/a_n$  playing a central role in the Fourier analytic theory (see e.g. [17], p. 98). As before, under suitable assumptions on  $x_k$ , the variables  $S_n/a_n$  and  $S_N/a_n$  are close to each other pointwise and thus we get a strong approximation of  $S_n/a_n$  with an infinitely divisible variable. In particular, it follows that the weak convergence of  $S_n/a_n$  along the whole sequence of integers or along a subsequence is equivalent to the convergence of the corresponding Lévy measures  $ndG(a_n x)$ . This is the well known 'generic' condition for weak convergence of i.i.d. sums, used e.g. to characterize domains of attraction and domains of partial attraction. It is typically proved by Fourier analytic methods and Lévy's pointwise approximation approach has considerable methodological interest.

For historical accuracy, one has to point out that strong approximation is nowhere mentioned in Lévy's paper and his sole interest was weak convergence of i.i.d. sums. Neither is the St. Petersburg paradox mentioned anywhere in [23], a curious fact, since Lévy was interested in the problem and discussed it in length in his book [22], pp. 122-133.<sup>2</sup> The great power of strong approximation in probability theory and statistics was not recognized until Strassen's paper [27] and strong approximation results based on the quantile transform such as those of Komlós, Major and Tusnády [19], or the strong approximation approach to domains of attraction and partial attraction due to Csörgő, Haeusler and Mason [10] give much better rates and have a wider scope of applications than Lévy's approach above. We refer to [6], [10], [26] and the references therein for history and further applications of the quantile transformation method.

As far as the distributional closeness of i.i.d. sums and their accompanying infinitely divisible laws is concerned, this has a wide literature, starting with Doeblin [12]. Lévy's observation that the accompanying infinitely divisible distribution of the partial sums  $S_n = \sum_{k=1}^n x_k$  is the same as the distribution of  $S_N$  with a Poisson distributed  $N$  has not gone unnoticed, see e.g. LeCam [20]. However, the usual path of estimation of the distributional closeness of  $S_n$  and  $S_N$  in the literature is not the direct path above, but, following Doeblin [12], it utilizes concentration function arguments (also going back to Lévy). For the uniform distance of the distributions of  $S_n$  and  $S_N$  Kolmogorov [18] obtained the bound  $Cn^{-1/5}$  with an absolute constant  $C > 0$ ; this has been improved by several authors, see Arak and Zaitsev [1] for a historical account and optimal results.

---

<sup>2</sup>See Csörgő and Simons [11], pp. 69–70.

### 3 Asymptotics via the quantile transform

In this section we discuss another argument in Lévy [23], standing somewhat apart from the main line of discussion of [23], but having profound consequences on the behavior of i.i.d. sums. On pp. 372–374 of [23] he makes the following remark.

**Classification of the terms of  $S_n$  in decreasing order.** *Let  $\xi_1, \xi_2, \dots, \xi_n$  denote the numbers  $x_1, \dots, x_n$  ordered in decreasing absolute values. The role played by the largest of the  $|x_\nu|$  in the previous discussions makes one think that there is interest in studying the random variables  $\xi_p$  and considering  $S_n$  as their sum.*

*Put  $y_p = F\{|\xi_p|\}$ . The numbers  $y_1, y_2, \dots, y_n$  are random variables chosen at random between 0 and 1 and then arranged in increasing order. Obviously,*

$$\mathcal{P}\{y < y_p < y + dy\} = pC_n^p y^{p-1}(1-y)^{n-p} dy$$

*from where we get easily, by using Euler integrals, that*

$$\mathcal{E}\{y_p\} = \frac{p}{n+1}, \quad \sigma^2\{y_p\} = \frac{p(n-p+1)}{(n+1)^2(n+2)}.$$

*If  $p \sim \alpha n$  as  $n \rightarrow \infty$ , the previous expressions are equivalent to  $\alpha$ , resp.  $\frac{\alpha(1-\alpha)}{n}$ , as one can deduce it immediately from Bernoulli's theorem. The case which is most interesting for us is when  $p$  is fixed and  $n \rightarrow \infty$ ; then we have*

$$\mathcal{E}\{y_p\} \sim \frac{p}{n}, \quad \sigma^2\{y_p\} \sim \frac{p}{n^2}$$

$$\mathcal{P}\{ny_p < \eta\} \sim \frac{1}{(p-1)!} \int_0^\eta \eta^{p-1} e^{-\eta} d\eta \quad (12)$$

*and thus, asymptotically,  $ny_p$  is a random variable with mean  $p$  and mean quadratic deviation  $\sqrt{p}$  and the normed difference  $\frac{ny_p - p}{\sqrt{p}}$  is asymptotically normal (provided  $p$  is very large). Further, between the  $y_p$  there is a positive correlation: if  $y_p$  is known, then  $y_{p+q}$  can be considered as the  $q$ -th of  $n-p$  variables, chosen between  $y_p$  and 1 and arranged in increasing order. The fact that each  $y_p$  differs little from its expectation permits us to neglect this correlation.*

*To fix the ideas, assume that the studied law is symmetric; each  $\xi_p$  is then a variable with modulus  $\rho_p$  known in terms of the  $y_p$  and with random sign. The sum  $S_n$  can then be written as*

$$\pm \varrho_1 \pm \varrho_2 + \dots \pm \varrho_p \pm \dots \quad (13)$$

*and here the number of terms grows with  $n$ , so that it can be considered as a series. If  $\frac{\varrho_p}{\varrho_1}$  has an order of magnitude independent of  $n$ ,  $S_n$  has the order of magnitude  $\varrho_1$  or larger, according as the probability of the convergence of*



the series is 1 or 0, i.e. according as the series  $\sum \varrho_k^2$  converges or not. It is in the case of divergence of this series when the law of large numbers applies.

Thus, if  $F(x) \sim x^{-\alpha}$ , we have (using  $p/n$  as an approximative value of  $y_p$ )

$$\varrho_p^\alpha \sim \frac{n}{p}, \quad \frac{\varrho_p}{\varrho_1} \sim p^{-\frac{1}{\alpha}},$$

and thus the law of large numbers applies if  $\sum p^{-\frac{2}{\alpha}}$  diverges, that is if  $\alpha \geq 2$  and only in this case. We can thus reprove well known results by an intuitive procedure which we found useful to mention, since the validity of law of large numbers is connected with the divergence of  $\sum \varrho_p^2$ .

This remark seems to be the first application of the quantile transform in probability theory, an argument used 40 years later with spectacular success to the asymptotic study of sums of independent random variables. To understand what is actually stated and proved here, one has to note that the phrase "the law of large numbers applies" for a sum of independent random variables means, in Lévy's terminology (see the top of p. 363 in [23]) something different, namely the uniform asymptotic negligibility of the terms of the sum compared with the sum. As Lévy proves in the remark above, this condition holds for sums of i.i.d. symmetric random variables with tails  $\sim cx^{-\alpha}$  iff  $\alpha \geq 2$ . But a more important consequence of the argument above is that for  $0 < \alpha < 2$  the  $p$ -th largest term of a sum of  $n$  i.i.d. symmetric random variables with tails  $\sim cx^{-\alpha}$  has the order of magnitude  $(n/p)^{1/\alpha}$ , i.e., the extremal terms of the sum have the same order of magnitude as the sum itself. This is a fundamental observation, the 'signature' property of i.i.d. sums with stable tails. Naturally,  $\rho_p$  depends also on  $n$  and relation (12) shows that for fixed  $p$  and  $n \rightarrow \infty$  the limit distribution of  $ny_p$  is Gamma( $p, 1$ ) (a fact also discovered much later). Thus using  $F^{-1}(x) \sim c_1 x^{-1/\alpha}$  ( $x \rightarrow 0$ ) and the connection between  $y_p$  and  $|\xi_p|$  we see that  $n^{-1/\alpha} \rho_p$  converges weakly, for fixed  $p$  and  $n \rightarrow \infty$ , to  $1/Z_p^{1/\alpha}$  where  $Z_p$  is a Gamma( $p, 1$ ) random variable. It is now tempting to divide in (13) by  $n^{1/\alpha}$ , let  $n \rightarrow \infty$  and conclude that a random variable having the stable limit distribution of  $S_n/n^{1/\alpha}$  has the infinite representation

$$\pm Z_1^{-1/\alpha} \pm Z_2^{-1/\alpha} \pm \dots \quad (14)$$

where  $Z_p$  is a Gamma( $p, 1$ ) variable. Lévy's justification of this transition to infinite series is a bit vague: "here the number of terms grows with  $n$ , so that it can be considered as a series" and his proof of the a.s. convergence of the infinite series is not complete either: for justifying the use of the Kolmogorov two-series criterion he remarks merely that the correlation between the terms of the sum can be neglected. The decomposition (14) was proved only in 1981 by LePage, Woodroffe and Zinn [21] with the crucial addition that the  $Z_p$ 's are the partial sums of a *single* i.i.d. sequence of exponential random variables with mean 1; this yields naturally the precise joint distribution of the terms of the expansion. The modern tool used in [21] is the

representation of the uniform ordered sample in the form  $(Z_1/Z_{n+1}, \dots, Z_n/Z_{n+1})$  (see e.g. [26], p. 335) and the proof of the a.s. convergence of the sum in Le Page, Woodroffe and Zinn [21] requires a delicate argument. As we will show in [3], this decomposition, adapted to semistable variables, leads to a strong approximation of St. Petersburg sums with a semistable Lévy process with a.s. error  $O(n^{1/2+\varepsilon})$  and an asymptotically normal remainder term, proving an unexpected central limit theorem in St. Petersburg theory.

**Acknowledgement.** The author thanks Professor Miklós Csörgő for his valuable comments.

## References

- [1] Arak, T. V. and Zaitsev, A. Yu.: Uniform limit theorems for sums of independent random variables. Proceedings of the Steklov Institute of Mathematics 174 (1986), 214 pp. American Mathematical Society 1988.
- [2] del Barrio, E., Janssen, A. and Pauly, M.: The  $m(n)$  out of  $k(n)$  bootstrap for partial sums of St. Petersburg type games. Electron. Commun. Probab. 18 (2013), 1–10.
- [3] Berkes, I.: Quantile decomposition and the St. Petersburg paradox. Preprint.
- [4] Berkes, I., Horváth, L. and Schauer, J.: Non-central limit theorems for random selections. Prob. Theory Rel. Fields 147 (2010), 449–479.
- [5] Bernoulli, D.: Specimen theoriae novae de mensura sortis. Comm. Acad. Sci. Imp. Petropolitanae 5 (1738), 175–192. German translation: A. Pringsheim: Versuch einer neuen Theorie der Wertbestimmung von Glücksfällen, Verlag von Duncker & Humblot, Leipzig, 1896.
- [6] Csörgő, M. and Révész, P.: Strong approximations in probability and statistics. Academic Press, 1981.
- [7] Csörgő, S. A probabilistic approach to domains of partial attraction. Adv. Applied Math. 11 (1990), 282–327.
- [8] Csörgő, S. Rates of merge in generalized St. Petersburg games. Acta Sci. Math. (Szeged), 68 (2002), 815–847.
- [9] Csörgő, S. and Dodunekova, R.: Limit theorems for the Petersburg game, in: Sums, trimmed sums and extremes (M. G. Hahn, D. M. Mason, D. C. Weiner, eds.) pp. 285–315. Progress in Probability, Vol. 23, Birkhäuser, Boston 1991.
- [10] Csörgő, S., Haeusler, E. and Mason, D.: A probabilistic approach to the asymptotic distribution of sums of independent, identically distributed random variables. Adv. in Appl. Math. 9 (1988), 259–333.

- [11] Csörgő, S., Simons, G.: A bibliography of the St. Petersburg paradox. Hungarian Academy of Sciences and University of Szeged, 2013.
- [12] Doeblin, W.: Sur les sommes d'un grand nombre de variables indépendantes. *Bull. Sci. Math.* 63 (1939), 23-32, 35-64.
- [13] Doeblin, W.: Sur l'ensemble de puissances d'une loi de probabilité. *Studia Math.* 9 (1940), 71-96.
- [14] Feller, W.: Note on the law of large numbers and "fair" games. *Ann. Math. Statist.* 16 (1945), 301-304.
- [15] Gnedenko, B.V. On the theory of domains of attraction of stable laws. *Uchenye Zapiski Moscov. Gos. Univ.* 30 (1939), 61-81 (In Russian).
- [16] Gnedenko, B.V. Some theorems on powers of distribution functions. *Uchenye Zapiski Moscov. Gos. Univ.* 45 (1940), 61-72 (In Russian).
- [17] Gnedenko, B.V. and Kolmogorov, A.N.: Limit distributions for sums of independent random variables. Addison-Wesley, 1954.
- [18] Kolmogorov, A.N.: Two uniform limit theorems for sums of independent random variables. *Teor. Veroyatnost. i Primenen.* 1 (1956), 426-436. (In Russian).
- [19] Komlós, J., Major, P. and Tusnády, G.: An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* 32 (1975), 111-131; II. 34 (1976), 33-58.
- [20] LeCam, L.: On the distribution of sums of independent random variables. *Bernoulli, Bayes, Laplace Anniversary Volume*, L. Le Cam and J. Neyman (eds.) pp. 179-202. Springer, 1965.
- [21] LePage, R., Woodroffe, M. and Zinn, J.: Convergence to a stable distribution via order statistics. *Annals of Probability*, 9 (1981), 624-632.
- [22] Lévy, P.: *Calcul des probabilités*. Gauthier-Villars, 1925.
- [23] Lévy, P.: Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchainées, *Journal de mathématiques pures et appliquées* (9) 14 (1935) 347-402.
- [24] Lévy, P.: *Théorie de l'addition des variables aléatoires*. Gauthier-Villars, 1937.
- [25] Martin-Löf, A.: A limit theorem which clarifies the 'Petersburg paradox'. *J. Appl. Probability* 22 (1985), 634-643.
- [26] Shorack, G. and Wellner, J.: *Empirical processes with applications to statistics*. Wiley, 1986.
- [27] Strassen, V.: An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. verw. Geb.* 3 (1964), 211-226.